# Molecular Identification and Characterization of Two Rubber Dandelion Amalgaviruses

*Archives of Virology – Annotated Sequence Record*

Humberto Debat[1*], Zinan Luo, Brian J. Iaffaldano, Xiaofeng Zhuang, Katrina Cornish [2]

[1] Instituto de Patología Vegetal, Centro de Investigaciones Agropecuarias, Instituto Nacional de Tecnología Agropecuaria (IPAVE-CIAP-INTA), 11 de setiembre 4755, Córdoba, Argentina, X5020ICA.

[2] Department of Horticulture and Crop Science, The Ohio State University, Ohio Agricultural Research and Development Center, Williams Hall, 1680 Madison Avenue, Wooster, OH 44691.

* Corresponding author:  Humberto Debat, debat.humberto@inta.gob.ar, Tel: +54 9 351 4973636, Fax: +54 9 351 4974330

**Abstract**

The *Amalgaviridae* family is composed of persistent viruses that share the genome architecture of *Totiviridae* and gene evolutionary resemblance to *Partitiviridae*. A single *Amalgavirus* genus has been assigned to this family, harboring only four recognized species, corresponding to plant infecting viruses with dsRNA monopartite genomes of ca. 3.4 kb. Here, we present the genomic identification and characterization of two novel *Amalgavirus* detected in Rubber dandelion (*Taraxacum kok-saghyz*). The sequenced isolates presented a 3,409 and 3,413 nt long genome, harbouring two partially overlapping ORFs encoding a putative coat protein and an RNA-dependent RNA polymerase (RdRP). Multiple independent RNAseq data suggest that the identified viruses have a differential distribution and low relative RNA levels in infected plants. Virus presence was not associated with any apparent symptoms on the plant host. We propose the name rubber dandelion latent virus 1 & 2 to the detected *Amalgavirus*.

**Annotated sequence record**

Natural rubber is an essential material to the manufacture of 50,000 different rubber and latex products. A steadily increasing demand cannot be met only by the rubber tree (*Hevea brasiliensis*). Viable alternative crops that could be established may supplement the demand, with carbon footprint savings, which is currently supported by diverse synthetic rubbers [1]. Rubber dandelion (*Taraxacum kok-saghyz*) is currently being developed as a sustainable source of natural rubber. Thus, a robust metabolomic, genomic, and transcriptomic characterization should advance in parallel to meet a biological landscape of this important natural resource [2]. In this direction, we searched the first publically-available RNA-Seq based *T. kok-saghyz* transcriptome, which was developed from pools of roots of genotypes with high and low rubber yields [3]. This transcriptome was produced from total RNA extracted from 6 month old root samples of *T. kok-saghyz* at The Ohio State University, and sequenced by Illumina Hiseq2000, obtaining 65,843,904 pair-end 100bp reads (NCBI SRA accession SRR5181667; 6.6 Gbp). The sequenced reads were quality evaluated using the FASTX-Toolkit, with a cut-off score of 30 (-q). The filtered reads then went through Trinity *de novo* assembly (version 2.2.0) using standard parameters. The NCBI SRA transcriptome assembly was subjected to bulk BLASTX-NCBI searches. Interestingly, two transcripts presented consistent sequence identity to the *Amalgavirus Southern tomato virus* [4] (50% identity at the aa level; E-value = 0.0) and *Blueberry latent virus* (49% identity at the aa level; E-value = 0.0). The corresponding transcripts were curated by iterative mapping of RNA reads, which gave a mean coverage support of 49.1

X and 77.7 X, respectively. The curated 3,409 nt and 3,413 nt long sequences were further explored in detail and designated tentatively isolate OH of rubber dandelion latent virus 1 & 2 (RdLV1 & RdLV2).

The RdLV1 genome presents a 143 nt 5´UTR, a 97 nt 3´UTR, and two partially overlapping ORFs on the positive strand (Figure 1.A). The predicted ORF1 encodes a 387 aa putative Coat protein (CP). The overlapping ORF2 encodes an 825 aa RdRP with a corresponding RNA_dep_RNAP domain (Pfam: pfam00680, E-value = 1.30e-07) at the 360-544 aa coordinates. Genome position 981 (_$) presents a putative "slippery" sequence of the form ACU_UUU_CGC suggesting a host ribosomal +1 frameshift signal that could induce the generation of a characteristic 1,055 aa, 120 kDa fusion *Amalgaviridae* protein (Figure 1.B). This slippery sequence is identical to the reported frameshifting signal of the *Amalgavirus Rhododendron virus A* [5]. The RdLV2 genome presents a 171 nt 5´UTR and a 100 nt 3´UTR (Figure 1.A). The predicted ORF1 encodes a 377 aa putative CP. The overlapping ORF2 encodes a 749 aa RdRP with a RNA_dep_RNAP domain (pfam00680, E-value = 2.99e-10) at the 304-473 aa coordinates. Genome position 946 (_$) presents a putative "slippery" sequence CAG_UUU_CGU that could induce the generation of a 1,046 aa, 118 kDa fusion protein (Figure 1.B). The UTR regions of RdLV1 & RdLV2 were A+U rich, as described for *Amalgavirus* [5], ranging from 53.1 % in the RdLV1 5´UTR to 61 % in the 3´UTR of RdLV2. The putative CP of RdLV1 & RdLV2 were subjected to 3D structure prediction with the EMBOSS 6.5.7 Tool Garnier and coiled coil determination by COILS with a MTIDK matrix. A comparison of these predictions to that of reported *Amalgavirus* (Figure 1.C) suggests that RdLV1 & RdLV2 present a typical α-helical central region with high probability of coiled coil as part of its tertiary structure, as is prevalent in *Amalgaviridae* [6]. It is worth mentioning that the predicted forms of potential slippery sequences of RdLV1 & RdLV2 are of the general form UUU_CGN, similar to the experimentally validated sequence of Influenza A virus [7]. Theoretically, the ribosome may stall on a slippery sequence, making a pause at a rare codon (such as CGN = R) for which scarce tRNAs might be available. This pause may lead to a movement forward of one nucleotide. Translation resolves on the advanced ribosome in the +1 frame (Figure 1.B). This phenomenon has been predicted to be widespread among most plant amalgaviruses [6]. RdLV1 & RdLV2 share a 55.9 % genome nt identity and a 49.5 % aa pairwise identity between their predicted RdRPs. Their proposed assignment as separate species is consistent with the species demarcation criteria for the genus *Amalgavirus* proposed by the ICTV, which specifies an amino acid sequence divergence of over 25% at the RdRPs. The structural highlights of RdLV1 & RdLV2 were compared to the ICTV recognized Amalgavirus species (Table 1). The predicted genome lengths and architectures, ORFs, UTRs, gene products, protein sizes, and general viral sequence cues are consistent with the proposed assignment of RdLV1 & RdLV2 to the *Amalgavirus* genus. The predicted RdRP of RdLV1 & RdLV2 were employed to glimpse some evolutionary insights of the identified viruses. Maximum-likelihood phylogenetic trees of RdLV1 & RdLV2, and reported amalgaviruses, in the context of related viral families were generated, based on MAFTT protein alignments obtained by the Geneious 8.1.9 platform (Biomatters Ltd.) and its FasTree plugin v1.0. The resulting trees evidently place RdLV1 & RdLV2 in a cluster of amalgaviruses, and more distantly related to new unclassified viruses and members of the *Partitiviridae* and *Totiviridae* families (Figure 2.A). The complete fusion protein (FP) of RdLV1 & RdLV2 was explored in sequence similarity among recognized *Amalgavirus* species (Figure 2.B), and with closely related species (Figure 2.C) using the Circolette tool [8], highlighting a stronger and broader link among the FP of RdLV1 & RdLV2 and reported amalgaviruses. Interestingly, sequence identity robustly falls beyond the *Amalgavirus* genus. Nevertheless, similarity with a species proposed to be a member of a new genus of fungi derived *Amalgaviridae*, the Z*ygosaccharomyces bailii virus Z* (ZbvZ) [9], is consistently low, supporting that both RdLV1 & RdLV2 are derived from plants. To confirm the presence of the identified viruses and explore their preliminary prevalence, we investigated five independent root total RNA samples of *Taraxacum kok-saghyz* which were further individually sequenced by Illumina Hiseq2000 generating over 291 million 100 bp pair end reads, ranging between 5.2 Gb to 6.7 Gb per sample (SRA accessions & names: SRR5181661, TK-R21; SRR5181662, TK-R18; SRR5181663, TK-R14; SRR5181664, TK-R10 and SRR5181665, TK-R9). Interestingly, the presence of the cognate viruses was confirmed in 4 of the 5 samples by iterative relaxed mapping of sequencing reads to the reference transcripts of RdLV1 & RdLV2 (Figure 2.D). Virus RNA levels varied among samples, ranging from 3.69 FPKM for RdLV1 in TK-R14, to 12.11 FPKM for RdLV2 in TK-R6. In addition, in the TK-R18 sample, only RdLV2 was found, and both viruses were absent in TK-R21, suggesting that RdLV presence is dynamic and that mixed infections, whilst common, are not necessary. *De novo* assembly of the raw RNA data and further identification of RdLV isolates on the diverse samples were carried out in order to address virus diversity. Sequence variants among samples were reduced, presenting a high degree of homogeneity. Overall identity among individuals ranged from 98.3% to 99.4%, which was roughly equivalent to the observed intra-individual identity which ranged between 99.2% and 99.5%. A consistent identity among isolates was reported for *Blueberry latent virus*, when 35 diverse cultivars were assessed and over 99% among isolates was observed [10]. Additionally, SNP were predicted (Figure 2.E), and 259 variants were identified among the CDS of RdLV1 & RdLV2; 78.37% of the polymorphisms involved the 3rd position of the predicted codon, suggesting a robust constraint to avoid amino acid changes and thus maintain structure and functional domains of the respective viruses. Recurrent attempts to transmit *Amalgavirus* via grafting and mechanical inoculation have failed. In addition, *Amalgavirus* are very efficiently transmitted vertically via seed (70–90%), and have been associated with symptomless infections in their respective hosts [4-5, 10]. The latter is consistent with our observations on tested rubber dandelions, which could not be linked with symptoms or altered phenotypes. Future studies should explore whether RdLV1 & RdLV2 share the biological properties of persistence and exclude potential horizontal transmission. To our knowledge, there are no reports of interspecific transmission of amalgaviruses, and transmission by potential vectors has not been conclusively ruled out. The identified RdLV1 & RdLV2 correspond to the first viruses associated with *Taraxacum kok-saghyz*. The molecular characterization of these prospective members of the *Amalgaviridae* family is a first step on the path to advance the understanding of the intriguing biology of these potential endophytes and their economically important plant host.

-Nucleotide sequence accession number: The genome sequences of Rubber dandelion latent virus 1 & 2 have been deposited in NCBI GenBank under accession no XXXX

**Acknowledgments**

This work was supported by the Center for Applied Plant Sciences (CAPS), and the College of Food Agricultural and Environmental Sciences, The Ohio State University, and USDA National Institute of Food and Agriculture (Hatch project 230837).

**References**

1. Cornish K (2017) Alternative Natural Rubber Crops: Why Should We Care? Technol Innov 18:244-255

2. Zhang Y, Iaffaldano BJ, Zhuang X, Cardina J, Cornish K (2017) Chloroplast genome resources and molecular markers differentiate rubber dandelion species from weedy relatives. BMC Plant Biol 17:34

3. Luo Z, Iaffaldano BJ, Zhuang X, Fresnedo-Ramírez J, Cornish K (2017) Analysis of the first Taraxacum kok-saghyz transcriptome reveals potential rubber yield-related SNPs. Sci Rep In review

4. Sabanadzovic S, Valverde RA, Brown JK, Martin RR, Tzanetakis IE (2009) Southern tomato virus: the link between the families Totiviridae and Partitiviridae. Virus Res 140:130–137

5. Sabanadzovic S, Ghanem-Sabanadzovic NA, Valverde RA (2010) A novel monopartite dsRNA virus from rhododendron. Arch Virol 155:1859-1863

6. Nibert ML, Pyle JD, Firth AE (2016) A +1 ribosomal frameshifting motif prevalent among plant amalgaviruses. Virology 498:201–208

7. Firth AE, Jagger BW, Wise HM, Nelson CC, Parsawar K, Wills NM, Napthine S, Taubenberger JK, Digard P, Atkins JF (2012) Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCC_UUU_CGU and is in the +1 direction. Open Biol 2:120109

8. Darzentas N (2010) Circoletto: visualizing sequence similarity with Circos. Bioinformatics 26:2620–2621

9. Depierreux D, Vong M, Nibert ML (2016) Nucleotide sequence of Zygosaccharomyces bailii virus Z: evidence for + 1 programmed ribosomal frameshifting and for assignment to family Amalgaviridae. Virus Res 217:115–124

10. Martin RR, Zhou J, Tzanetakis IE (2011) Blueberry latent virus: an amalgam of the Partitiviridae and Totiviridae. Virus Res 155:175–180

**Figure Legends**

**Figure 1.** (**A**) Rubber dandelion latent virus 1 & 2 (RdLV1 & RdLV2) linear monopartite dsRNA genome are 3,409 & 3,413 nt long, arranging a translation strategy based in two partially overlapping ORFs. The RdLV1 genome presents a 143 nt 5´UTR and a 97 nt 3´UTR. The predicted ORF1 encodes a 387 aa putative Coat protein. The overlapping ORF2 encodes a 825 aa RNA dependent RNA Polymerase. Genome position 981 (_$) presents a putative “slippery” sequence that could induce the generation of a 120 kDa fusion protein. The RdLV2 genome presents a 171 nt 5´UTR and a 97 nt 3´UTR. The predicted ORF1 encodes a 377 aa putative Coat protein. The overlapping ORF2 encodes a 749 aa RdRP. Genome position 946 (_$) presents a putative “slippery” sequence that could induce the generation of a 118 kDa FP. (**B**) Potential programmed ribosomal frameshifting of RdLV1 & 2. The RdLV1 ACU_UUU_CGC motif and RdLV2 CAG_UUU_CGU motif, of the general form UUU_CGN, are +1 ribosomal frameshifting motif prevalent among most plant amalgaviruses. (**C**) 3D structure prediction of the corresponding Coat proteins of RdLV1 & 2 and of reported amalgaviruses, assessed with the EMBOSS 6.5.7 tool Garnier represented on top, and

coiled coil determination by COILS with a MTIDK matrix as a line graphs. Regions of high coiled coil probability are constrained to the typical α-helical central region of the CPs.

**Figure 2.** (**A**) Maximum-likelihood phylogenic tree of the RdRP predicted protein of reported amalgaviruses in the context of related viral families based on a MAFTT multiple alignments. Numbers at the nodes indicate percentage of bootstrap consensus support values obtained for 1000 replicates. Sequence similarity levels of amalgaviruses Fusion Proteins among the *Amalgavirus* genus (**B**) and RdRP proteins of related viruses (**C**) expressed as Circoletto diagrams. FPs or RdRPs are depicted clockwise, and sequence similarity is visualized from blue to red ribbons representing low-to-high sequence identity. (**D**) Virus RNA levels expressed as FPKM of NGS sequenced rubber dandelion total RNA root samples. Values for RdLV1 are depicted in blue columns and values for RdLV2 in orange columns. (**E**) RNAseq based read mapping graphs of RdLV1 and RdLV2 with the 6 combined RNA libraries. Tracks from top to bottom represent coverage per base, sequence identity from red to green (higher), and SNP prediction. GenBank accession numbers and abreviations for the respective viruses are *Southern tomato virus* (STV, NC_011591), *Rhododendron virus A* (RV-A, NC_014481), *Blueberry latent virus* (BBLV, NC_014593), *Vicia cryptic virus M* (VCV-M, EU371896), *Hubei partiti-like virus 59* (Hplv, APG78262), *Beihai barnacle virus 14* (Bbv14, APG78182), *Zygosaccharomyces bailii virus Z* (ZbvZ, KU200450), *Colletotrichum higginsianum dsRNA virus 1* (Chv1, NC_028242), *Heterobasidion partitivirus P* (HpP, AAK52739), *Radish partitivirus* (AY748911), *Vicia cryptic virus* (VCV, EF173396), *Saccharomyces cerevisiae virus L-A* (ScV-LA, NC_003745), *Penicillium stoloniferum virus S* (PsvS, NC_007539), *Aspergillus ochraceous virus* (AoV, EU118277), *Cryptosporidium parvum virus 1* (CpV1, CPU95995), *Pepper cryptic virus 1* (PCV1, JN117276), *Trichomonas vaginalis virus* (TvV, NC_003824), *Fig cryptic virus* (FCV, NC_015494), *Atkinsonella hypoxylon virus* (NP_604475). α: *Alphapartitivirus* genus, β: *Betapartitivirus* genus, γ: *Gammapartitivirus* genus, δ: *Deltapartitivirus* genus, C: *Cryspovirus* genus.

| *Amalgavirus* | GS | 5´U | OR1 | OR | OR2 | 3´U | SLPs | SLPp | FP | RP | CP | RPi | CPi | GSi | Accession n. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Rubber dandelion latent virus 1** | 3,409 | 143 | 1,164 | 473 | 2,575 | 97 | ACU_UUU_CGC | 981 | 1,055 | 825 | 387 | ----- | ----- | ----- | XXXXXXX |
| **Rubber dandelion latent virus 2** | 3,413 | 171 | 1,134 | 242 | 2,250 | 100 | CAG_UUU_CGU | 946 | 1,046 | 749 | 377 | 49.5 | 21.5 | 55.9 | XXXXXXX |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Southern tomato virus* | 3,437 | 137 | 1,134 | 233 | 2,289 | 110 | CUU_AGG_CGU | 983 | 1,063 | 763 | 378 | 49.5 | 22.0 | 55.7 | NC_011591 |
| *Blueberry latent virus* | 3,431 | 166 | 1,128 | 359 | 2,397 | 99 | UCU_UUU_CGU | 979 | 1,055 | 799 | 376 | 46.2 | 19.5 | 54.8 | NC_014593 |
| *Rhododendron virus A* | 3,427 | 94 | 1,215 | 405 | 2,424 | 47 | ACU_UUU_CGC | 1,181 | 1,078 | 808 | 405 | 48.0 | 22.1 | 54.4 | NC_014481 |
| *Vicia cryptic virus M* | 3,434 | 142 | 1,185 | 287 | 2,277 | 117 | ACU_UUU_CGU | 1,089 | 1,058 | 759 | 395 | 47.3 | 19.7 | 54.8 | EU371896 |

**Table 1** Diverse structural highlights of RdLV 1 & 2 in comparison with ICTV recognized *Amalgaviridae* species. GS: Genome size (nt), 5´U: 5´UTR length (nt), OR1: ORF 1 length (nt), OR: Overlapping region (nt), OR2: ORF 2 length (nt), 3´U: 3´UTR length (nt), SLPs: Slippery sequence, SLPp: SLP position, FP: Fusion protein length (aa), RP: RdRP (aa), CP: Putative Coat protein length (aa), RPi, CPi, GSi: RdRP, CP, and complete genome sequence identity of the corresponding amalgavirus in relation to RdLV1.

**Sequences**

Considering that the GenBank deposited sequences would be released upon acceptance of the manuscript, we provide here, in plain FASTA format, the complete sequences of Rubber dandelion latent virus 1 & 2

```
> Rubber dandelion latent virus 1
GCCAUAUUUUGCUUACUCAACUGGUGUGUCGUGCCCUAUCGCAGUGUGUGCUUGCGCCUG
CUGUUUGCCCCCCUGUUUACUUUCUUCUUUAUUAUUCUCUUGCAACCCCCUGUGUUUGUU
GUUUCUUGUGUGUAGCUUACAGGAUGUCAGGUUCUGGUGCUGCUUCUGGUUCCAAUGUC
CAUGCCGCUCGCCGUUCUGUUGAUUACGAGUCUAUCCUGGGGGAGCAGCUCGCCAACCUU
GCCCCUGAAGCCUUCCCUGUCUCAGACUGGACCAUCCCAAACAUCACCAAGUCCUUCCUG
ACGGUGCCAAAGUUCAUCGACACGAUCAAGGUUCUCACUGCCUGCGGCGACCCCGUCCUC
AUCAGAAGAAUCGCUGCCCAGGCCAUCACCCGUCAAGCAUGGGAGUCCAACACCACGUGC
ACCAUUCCCCAGAUGUUCAAGUUCUGCACGUGGCUGCGCACUCCCCAGGGUACUGAGCUG
AUCAACGACCUGCGCCGUACGCGCAAUCUGGAGAAGAAGACUGUGGGUGAACAGUCUAU
UGAAGAUGUUGGGUACGUGGGCGCCUUGGAGCAGAUGUUGACUGAUCGAGCCCUGGAAA
UCAAGCUUACCAGGGCCGAUUAUGACAACAGGCUGGCUGAGGCGAGGAGGCAGAUUGUC
UUGUUGGAGAUGGAGAAGGAGUCCAAGUUGAAGAGGAUCGAUGAGCUCUUCCAGCCUGC
GUCAUUCUAUGUUCCUUUGGAUGACAUGGAUCUGGGAAUGCAGUGUUAUGAGCUCUACC
AGAAAGAGUGCGCCGCGCUCGGCAAAGACGAAGCCCCGUUUGACGAGCAUUUGAUGGAG
GAUGUGCGGGCCACCUAUCAGAAUCAGGCCUUGGCUAAGCACAAGGCGGAGUUUGUUCG
GGAUGAGGAUAGGAGAAAUGCCAUCAAGUACUGGGUGGAGAAGAAGAUUCUGGAGUUGG
AUGGGAGAGGCGACCGUCGCCUCGCGCAGACUUUUCGCUCCUACCUUGCUGCACAAGGUG
GGAGACUGGAUGAUGAAGUUCGAACUGCCAUUAAGACACGACUUGGUGAGAAAGAUGGU
UCUGGGCCAUCCACAGAGGGACCUACGCCGGAGGGGGAACUUUUCGGUGACUUUGAGAG
AGUUUCCCCCGACCGUCGGGGAGCGGAGGACGAGGGGAUUCAGGUGGAACCCCCCGCUGU
CGAAGUUGUCGGAGCGGACCCCGGGGGUCACAGGUACGCUACAAGGAAACGCCCGCCUCC
ACGUGGUAGUCCCCCCAAGGGCAGGGGAAGGGGGAAGCAUCCCCACGGCAAGAAGUAAGU
UUGAAUCCAAAGUGCGCAAGGUUAUCGGGGGGGGUGAGAUGAGGGGUUGGCGGGAAGAC
ACGGCGAAAUACCGGGGAGGGGGUAACUUCUGUGACGCUAUCAAACUGCUUGCAGAUGC
```

GUCCGUGAAGGUUCCUGGCCGUACUCUUCGGGAUUGUUAUACUGUGGAUACUGCCCGCCU
UGUUCUCAAGCUACCUUGCGGACUUGGAGUACCUCGGGGGCCCGAGAGUGUAAUAAUGA
AGAACUUUAAUAAUGAAGCGACGGCGGGUCCGUGUAUGCGGGCUUUUGGAAUAAGGAGG
AAGUAUGGUCUUAAGAGGGGGAUGGAGGAGUUUGCGUGGAGUUGCCUGGAUGCGUAUGC
CCUGGGGGGGCGUCUCGAACGGAGCUUGCCCUAUGUUGCUGCUAGGGUGGGCUUUCGGAC
CAAGUUGCUCGAACAGAAAGAAGCCAUGAGAAAGAUCGCCGAUGGGAAACCUCUGGGCA
GAGCCGUGAUGAUGCUUGAUACGCACGAGCAGGUUUUCUCUUCUGCUCUCUACAACGUAC
UAAGCGGUCUGACUAAUCGGGCGAGACACACGCGGGAAAGUGGUUUUCGUAAUACCACU
AUACGCGCCUCCUCAGACUGGGCAAUUCUUUGGGAGGAGGUGCGCGAUGCUUCUGCUGUG
GUUGAGCUCGAUUGGAAGAAAUUUGACAGGGAGAGGCCCGCUGACGACAUCCAGUUCAU
GAUCGAGGUAAUUUGUUCCUGUUUUGAGCCCAAGGACGUUUAUGAGGAGAGGUUAUUAG
AGGCGCAUAGGAUAAUGUUGAAUCGGUCGUUGAUUGAGAGGCCCCUCAUCACUGAUGAU
GGCGGGGUCUUCACAAUCGAGGGCAUGGUGCCGAGUGGUUCUUUGUGGACCGGCUGGCU
CGACACUGCCCUCAAUAUUCUUUAUAUAACGGCGGUUCUUCGUUUCCUGAACUUUGACUA
UAAUGAUGCCGUCCCGAAAUGCGCCGGUGAUGAUAACCUCACUCUGUUUUAUACUGACGU
UAAUGAUGCUGUGCUUAAUAGAAUUAAGGUACUACUCAAUGAGUGGUUCCUGGCUGGCA
UUGAGGAUGAGGAGUUCUUGAUCCAUAGGCCGCCCUUCCAUGUUGGGCGUGUCCAGGCA
GUUUUUCCUCCUGGUACUGAUCUCUCCCAGGGCACAUCGAAAAUGUUGGACCAAGCAGAA
UGGAUUCCGAUCGAGGAGGAAAUGAUAAUAGACGAGCCAGCCGGCCUGUCACACAGGUG
GAAGUAUACAUUCGAUGGGAAGCCGAAAUUUCUGUCCUGCUAUUGGGAUAGAUUUGGUA
ACCCAAUCAGGCCCUCAUAUAUCAAUCUGGAGAAGUUGCUCUGGCCAGAAGGUAUACAUG
CCACUAUUGAUGACUACGAGGCGGCGGUCAUCAGCAUGGUCGUUGAUAAUCCUUUCAACC
ACCACAACGUCAACCACAUGAUGCAUCGUUACUGUAUCAUCCAACAAGUCAAACGCAUUG
CAGUAACAGGUAUUAAGCCCGAGGACGUCCUGACGUUAUGUAAGUUCAAGGGCGGGGAG
GACGAACCUGUAGUCUUCCCCAUGGUUGCGGAGUGGCGGCGUGUCGAUGGAUGGGUCGA
UAUGGAGAAGCUUCCAAACAUCAAAGGGUACGUGGAUCAGUUUAAGAGCUUCGUCCAGG
GUGUGUCCUCCUUGUACACACGCUCGCCCAGGGGUGGGCUGGAUGCCUGGCGGUUUAUGG
ACAUAAUUAGAGGGUUCGAUCAUCUCGCGGAAGGUCAGUUCGGUAAUGACCUCGAUGAC
UGGGUCUCGUUUCUCAAGAACCAUCCCGUGUCAAGGUAUCUAAAGCCAACACGUGGGAAU
CGAGAGGUAGAGCAGGCGAAGGAAAUGUCCCUGGAGACACAGCAGAGAUUUAACAGGUU
UAGGAUGGCGCUCCACCCCCUCAGAACAUCAAAUUUCUUCGACAGCAUGGAGAGUUACGC
UAGAUGGAUCUCAGAGUCGCUUAGAAACCGUGGGUCUAUGUAAUCAUUUGUUUGUUUCC
UGUAUUGUAAUAUGCAUUAUCUUUAAUAUUAACGUAUGUGAGCCUCCAUCGUGCGAACC
CCCCGAGCGCAUGUAUGGCGGG

> Rubber dandelion latent virus 2
GCUGAACUGAAGUUCUGCGUAUUUCUCCUGCUGUGUAUUGCCCGCUGACUUCUUGCUUUC
UUGUGUUUAUUCGCAGGUUUCAUAUCGUUCAGGUUCCAUCUCUUCUUGCAUUAUUAUUU
UCACUAACGCAACCUCUACUUGUUUUAGUGUGAAUUGUGUGCAGAUCUCAUCAUGGAAU
UCGUUCUCGAGAAGACUCCUGCUGAAGAACAAGCAAUGCUCGUCGAGAAAGCCGCUCCUC
UCAUCGCCUUUCACUUCCCUGCCUCCAUCUUCACUCGCGAUGCUGCCAUCGACGCGGGUU
ACACCUUUAAGAGCUUUCUCAAGCAUGUCACUUCCGUUGCUCGUUAUGCGGAGAGUGACC
AACUGCGAGCCAUCUGCAGGCUCGGGAUCAAGCAUGACAUCUUCGAGCUUCAUCAGGAGU
GUUCUAUCGAUCAGUUUGUCCGCUUCUCGGACUUUCUGAAGUCAAAGGAAGGGCAGGCG
GCCCUUCAGGGUGUGGCCAUUCAGACUAAGUUACAGAAACGGGCCGGCACCACGUUCACC
CCCAAGGAUGUAGCCCUGGAGCAGAUUUUCUCGAUUAUGCGCCAGGAUUUCCAUGCUGCG
AUGAAGGAGGAGGGGGCGGCUUUUCAGAAAGUGUUGGCCGAGUUGCGUCUCCAGAUCAA
GAAAGUUGAGAAAGAGUGGGAAGACCGUCAAGCGGAGAUCCGGGCCGCUUUCAAUCCUG
UGUCUGUCUACCAGGAGCCCAGUGAGAAGGAUAUAGGCGUGGAGGCCUAUGCUGCGUAU
GAGAAGGAGGCGGGUCUCAAGAAUAUGGUGGCCAAGCCUAAAGCUUCCGGGGGGCUCGA
GUACGCCAUCCAGAAUUAUGGCCCCCAGAUCUCCCGGACACGUGUUGUUGAGUUCGCGAA
CUUGCCCGAGCAUCAUGAGGGCUUCUUCGGGUAUAUGAAGCAGCGGGUGCUUCAGUUUC
GUACCCAGUUCGACACCAAGCAGGAAAAGGCAUUCGUCAAUCUCGUGGUUGCCGCGGGUG

GAGGAGAAAUUGUUAUCACUCCCGAAGACGAGAAGGUGGGCCCUGUCCUCUAUGUGCCCC
GUGGGGUUACCUCCUACCCCAGAGCUGAGGCCGUCUUGCUUCCCCCUGUCCGACCUAAUA
AGCGAGGAGGUGAUGAACAGGAAGGCCCAAGUCGGCGUAAGGGUCGGAAACCGUCUGGU
AAAGGGCCUAAGACAAGGUCGCAGCAUAGCAAAGACGCUGGCGAGCACUCGGAUUCAGG
UUCUACAGCCCAUAAGGGAGGUGAAGAUGAGAGCCAUCCCCACGGCAAGAAGUAAGUGG
GAGUCGGCGCUCAGGCGGAUUAUUGGCGGUGGUUCUAUGCGUUCAUGGCAGAAGGAUAG
UGAUAUGAUGCGAGGUGGUGGAGAUGUUGCGGACGCUGUCCUUUUGCUUGGAACUGCCG
UCGAUGAGUCGCCUGAGCGCCUGUUGCGGGGGCUUUUCAGCCCGGAGUUUGCGAGGAAG
GUUCUGCUACUUCCUUCCGGGUUGGAAGUCCCCGACGGGUUGGAGAUGUGUAGGAUGAA
GAACUUCAAUGAGGAAGCGACAGCUGGGCCGUUUCUGAGAGCGUUCGGAGUUAAAGGGA
AGUAUGGUUUGAAGGGGGUUUUGGAGGAGGAGAUGUGGUGGUAUUAUGAUGCAUUUGCU
CGUGGGGAAUUGACUCCCGAGCAGAUGCCUCAUUUUGGGGCGAGAGUCGGUUUCCGGAG
CAAGUUGCUAGCCGGGAAGAAGUUUAAUGAGAAGGUGGCGGCUGGGGAGCCCUUGGGGA
GGGCUGUGAUGAUGUUGGAUGCGCUGGAGCAGGCUGCCUCGAGUCCGUUGUACAAUGUG
AUUUCUGCCUACACUUCUCGUAGGCGGCUUGAAGCAGCUUGCGGGUUCAAGAAUGGGGU
GAUAAAGGCCAGCUCUGAUUGGCCGAAGGUUUGGGAGGAGGUGAAGAAGGCGCAGGUAA
UUGUGGAGUUAGAUUGGAAGAAGUUUGACCGGGAGCGUCCUGCGGAAGACAUUGAUUUC
AUUAUCGAUGUCGUGAUUGGUUGUUUCUCCCCGCAGAGCUCCCGGGAAAGGCGAUUGCU
GGAGGGGUACAGGUUGAUGAUGCGCCGGGCCCUGGUGGAGAGGUUGGUUAUAAUGGAUG
AUGGGGGGGUGUUCGGGAUUGAUGGAAUGGUGCCGUCCGGAUCUCUCUGGACGGGAUGG
UUGGAUACGGCGCUUAACAUUCUUUAUAUACGGGCAGCGUGUGUGGAGGCGGGAUGUGC
GCCCCUAUCCUUCAGUCCUAUGUGCGCUGGGGAUGAUAACCUGACGCUGUUCUACAGUGA
CAGGGAGGAUUCAGUGUUGCUGAGGAUUAAGGGGCUCUUGAACAGGUGGUUCCGCGCGG
GCAUUGGAGACGACGACUUCAUAAUUCAUCGGCCGCCCUUCCAUGUUAUCAAGCAGCAGG
CGGUAUUCCCACCAGGAACCGAUCUUUCUCACGGCACUUCGGCCAUAAUUCAUUUGGCUA
AGUGGGUGGAGUUCGAUGGCGAGCUCGAGAUUGAUUUGGAUUCUGGGAAGUCUCAUAGA
UGGGAAUACGUAUUUAAGGGUAAGCCAAAGUUCCUUUCGAACUAUUGGCUUCUGGAGGG
GCAGCCGAUCAGGCCGACGACAGACAACCUCGAGAAGCUUCUCUGGCCGGAAGGGAUUCA
UGACGACCUAGAUGAUUAUCAGGCGGCCUUAACAGCAAUGGUCGUUGAUAAUGUCUGGA
ACCAUCAUUUGGUUAACCAUAUGAUGAUGAGGUAUGUGAUCAUCCAGCAGUUACGUCGC
AUGAUGUUUAUCCGGGGUGCAGAUGAUGACAUCUGCUUCUGGGCGACGUUAAGGGAGAA
GGGGGGUGGGGUAAUACCCUAUCCGCAGGUAGCACCAUGGCGAAGAGGAUCAAGUCAGA
AGCGGAUGGAAGAUUAUCCUGAAACUCGCAGUUGGAUCGAGGACUUCUCCUCAUUUGUU
CAGGGAGUGACGUCGUUAUACUCCCGAGAUUGUGAGGGGGGCAUUGACAGUUGGCAGUU
CAUGAAAAUCAUCCGCUGUGAGGGCCAUGUUGGGGAGGGGCAGUACGGCAAUGAUCUCA
CUCGAUGGCUGACGUUUCUCUCUGAGAAUCCCUGCACCAAGUAUCUCAAGGCGGUGCGGG
GCUUGAGGAGGGGGCCCGUGGCCAAGGUGGGGGAGCCGGAGCAUCUCCAAGCUGUUGGU
GACGCAUUCGGAGUUUUGAGGGAGGCUCUCUUGUCCGGUCGCAUGGAAGGGGAGGGAAG
UUUUGCAAUUUGGGUUUCCGAUAGGUUAAUAGGAAACAAUGUAAAUGUUUAGCAUGUAG
UUUGUUUUCUGUAUUUCCUUCAUGUAAUGUUAAUAUUUAAUGGUAAUAUAUUUUGUUGG
AGUGUGCGCGGCCCAUGCAUUCAGCGCACACCCG